

PROTEIN STRUCTURE BY FOURIER TRANSFORM INFRARED SPECTROSCOPY:  
SECOND DERIVATIVE SPECTRA

Heino Susi and D. Michael Byler

Eastern Regional Research Center  
Agricultural Research Service, U. S. Department of Agriculture  
600 East Mermaid Lane, Philadelphia, PA 19118

Received July 22, 1983

Second derivative Fourier transform infrared spectra of the proteins ribonuclease A, hemoglobin, and  $\beta$ -lactoglobulin A (native and denatured) have been obtained in deuterium oxide solution from 1350 to 1800  $\text{cm}^{-1}$ . The relationship of the original spectra to their second derivatives is briefly discussed. In the second derivative spectra, clearly resolved peaks are observed which can be associated with the  $\alpha$ -helix,  $\beta$ -strands, and turns. No protein spectra with such resolution have heretofore been reported. Tentative assignments are proposed, and the observed peaks are related to the secondary structure of the proteins studied. The data appear to present the first direct spectroscopic evidence of turns in a native protein.

Infrared absorption spectroscopy in the 1400 to 1700  $\text{cm}^{-1}$  spectral region has been extensively used to study the secondary structure of polypeptides and proteins in the solid state (1) and in deuterium oxide solution (2-4). Deuterium oxide is usually employed instead of water because of its greater transparency in the region of interest. More recently, related Raman studies have been carried out in great detail (5,6). In  $\text{D}_2\text{O}$  solution the most useful absorption band for secondary structure studies is the amide I' band at 1620-1690  $\text{cm}^{-1}$ , which involves essentially C=O stretching vibrations of peptide groups (1,2). Each type of substructure, such as the  $\alpha$ -helix,  $\beta$ -strands, and the various kinds of "turns" (7), gives rise, in principle, to different C=O stretching bands. An analysis of the intensities and frequencies of these component bands in the infrared (4) or Raman (5,6) spectra yields qualitative, as well as semiquantitative, information about the secondary structure of the specimen. Unfortunately, even powerful spectrophotometers cannot resolve these characteristic spectral components because of their inherently broad line shapes (large half widths). Band resolution by curve-fitting techniques (4,6),

is also hampered by lack of information regarding component bands both for infrared (4) and Raman (6) spectra. The second derivative of the original spectra offers a direct way to identify the peak frequencies of characteristic components and thus permits much more detailed qualitative and, eventually, quantitative studies. This approach has been successfully used on commercial polymers such as polyvinyl chloride (8) but not, to our knowledge, on polypeptides or biopolymers.

#### MATERIALS AND METHODS

Bovine hemoglobin (H2500) and bovine ribonuclease A (R4875) were obtained from the Sigma Chemical Company\*. Bovine  $\beta$ -lactoglobulin A was supplied by Dr. Harold M. Farrell, Jr. of this Research Center. Fourier transform infrared spectra were obtained on a Nicolet 7199 FTIR spectrophotometer equipped with a Hg/Cd/Te detector, a Globar source, and a Ge/KBr beamsplitter. For each spectrum, a 4000 scan interferogram was collected at 2  $\text{cm}^{-1}$  resolution; prior to Fourier transformation, the interferogram was 1X zero-filled and apodized with the Happ-Genzel function. Second derivative spectra were calculated with the Nicolet software parameter DR2. (DR2 uses the function:  $A''(n) = A(n+1) - 2A(n) + A(n-1)$ , where  $A(n)$  and  $A''(n)$  are the intensities in absorbance units at data point  $n$  of the original spectrum and of the second derivative spectrum, respectively. Note that DR2 gives the second derivative without any smoothing.) Smoothing was accomplished prior to plotting with a nine-point Savitsky-Golay function (9). The protein solutions in  $\text{D}_2\text{O}$  were 5% W/V; the pD was adjusted by small additions of DCl or NaOD. (To avoid unwanted spectral complications, no buffers were used.) The cells used had  $\text{CaF}_2$  windows and 0.075 mm path length. Spectra of pure  $\text{D}_2\text{O}$  at the proper pD were subtracted from the protein solution spectra prior to obtaining the second derivatives. To assure complete isotopic exchange, solutions were prepared in advance and held at room temperature for 48 hours. (Spectra of partially deuterated proteins would be very difficult to interpret.) The disappearance of the amide II band in the 1550  $\text{cm}^{-1}$  region (2) indicates that the exchange was complete.

#### RESULTS AND DISCUSSION

Figure 1A presents the infrared spectrum of ribonuclease A; figure 1B, the second derivative of this spectrum; and figure 1C, the smoothed second derivative (see Materials and Methods section). A few words are in order at this point about second derivatives of infrared spectra in general. The intrinsic shape of a single infrared absorption line is approximated by a Lorentzian function (10), i.e.,

$$A = (s/\pi)/(s^2 + v^2),$$

\*Reference to brand or firm name does not constitute endorsement by the U. S. Department of Agriculture over others of a similar nature not mentioned.

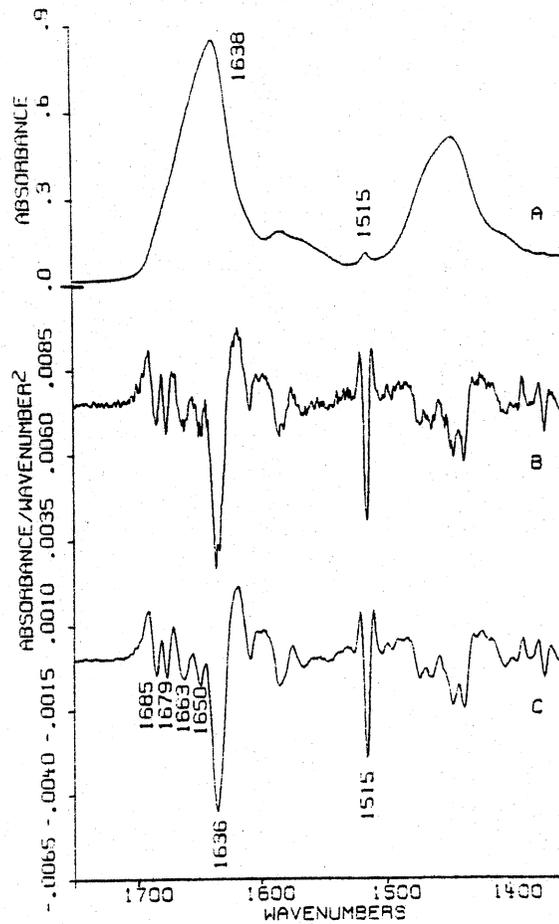


Fig. 1 (A) FTIR spectrum of ribonuclease A; (B) second derivative spectrum, unsmoothed; (C) second derivative spectrum, smoothed. 5% W/V in D<sub>2</sub>O; pathlength = 0.075 mm; pD 7.

where  $A$  is the absorbance,  $2s$  is the width at half height and  $\nu$  is the frequency with reference to the line center. The second derivative is (8,10)

$$A'' = -(1/\pi s) [2a(1 - 3av^2)/(1 + av^2)^3]$$

where  $a = 1/s^2$ . The peak frequency is identical with the original peak frequency. The half width of the second derivative,  $s^{II}$ , is related to the half width of the original line (10) by  $s^{II} = (1/2.7)s$ , and the peak intensity of the second derivative,  $I^{II}$ , to that of the original intensity by

$$I^{II} = -2I/(2.7s^{II})^2.$$

The peak intensity of the second derivative is thus proportional to the original peak intensity and inversely proportional to the square of the half

width. For real spectra with overlapping lines which might differ from the Lorentzian shape, the relationships are more complex. The above formulae do, nonetheless, provide a good approximate method for interpreting second derivative infrared spectra (8,10,11). We concentrate here mainly on the interpretation of the amide I' region ( $1620-1700\text{ cm}^{-1}$ ) which is related to the secondary structure of the proteins (1-6).

Figure 1C gives the smoothed second derivative spectrum of ribonuclease A. This protein has an approximate  $\alpha$ -helix content of 22% and a  $\beta$ -strand content of 46%, based on the classification by Levitt and Greer (12). By comparison with earlier studies (1-4), the  $1636\text{ cm}^{-1}$  peak can be associated with the  $\beta$ -segments and the much weaker  $1650\text{ cm}^{-1}$  band, with  $\alpha$ -helical segments. The weak peaks between  $1660$  and  $1700\text{ cm}^{-1}$  evidently correspond to the second  $\beta$ -structure band (1-4) and to turns (13,14). Comparison with the spectra of ribonuclease S (16), which has a very high  $\beta$ -structure content and relatively few turns (12), suggests that the  $1679\text{ cm}^{-1}$  band can be associated with the  $\beta$ -strands, while the  $1663$  and  $1685\text{ cm}^{-1}$  bands are due to turns.

In the second derivative spectrum of hemoglobin (Figure 2A) there is a single strong peak in this region, obviously representing the  $\alpha$ -helix. Hemoglobin is about 80%  $\alpha$ -helical and contains no  $\beta$ -structure (7). The weak peaks at ca.  $1638$  and  $1675\text{ cm}^{-1}$  can therefore be assigned to turns. The frequencies of the weak peaks are close to infrared frequencies reported for model turns by Kawai and Fasman (13), namely,  $1635$  and  $1695\text{ cm}^{-1}$  for one model, and  $1638$  and  $1687\text{ cm}^{-1}$  for another. The experimental values are also in reasonable accord with theoretical calculations by Bandekar and Krimm (14). If correct, these results would represent the first instance of observed vibrational bands which are assignable to "turns" in actual proteins. (In circular dichroism spectra, the most common technique for protein secondary structure studies, no specific features are observed for turns. Nor have any specific bands for turns been observed in Raman spectra of actual proteins.)

Figure 2B gives the second derivative spectrum of native  $\beta$ -lactoglobulin A. The crystal structure of this protein at high resolution has not been deter-

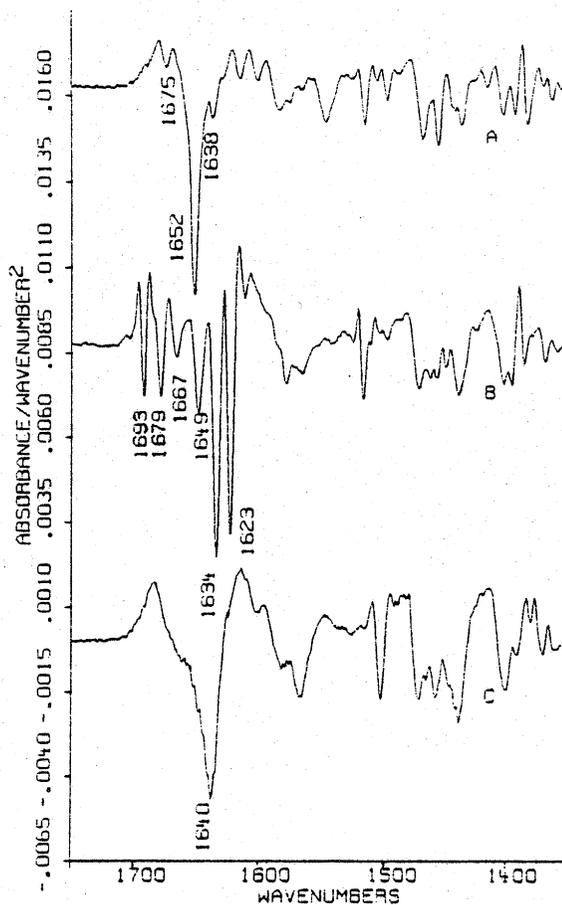


Fig. 2 Smoothed second derivative FTIR spectra: (A) hemoglobin, pD 7; (B) native  $\beta$ -lactoglobulin A, pD 7; (C) denatured  $\beta$ -lactoglobulin, pD 13. 5% W/V in D<sub>2</sub>O; pathlength = 0.075 mm.

mined, but circular dichroism (15) and infrared studies (4) suggest an  $\alpha$ -helix content of about 20% and a  $\beta$ -structure content of 45-50%. The 1650  $\text{cm}^{-1}$  (weak) and 1635  $\text{cm}^{-1}$  (strong) peaks are evidently associated with these substructure classes; the 1667 to 1693  $\text{cm}^{-1}$  bands, with the second  $\beta$ -structure band and with turns, as in ribonuclease A. It is interesting to note that the 1623  $\text{cm}^{-1}$  peak is absent in ribonuclease A but present in concanavalin A (16) which also has a very high  $\beta$ -content (12). This peak probably represents a variation of the  $\beta$ -structure (or a special type of turn) present in concanavalin A and  $\beta$ -lactoglobulin A, but not in ribonuclease A. Further study is evidently required. Figure 2C gives the second derivative spectrum

of denatured  $\beta$ -lactoglobulin A, which is assumed to be in a "random form" (2). Two observations are pertinent: (a) all sharp peaks associated with helix, extended chain, and turns have disappeared; (b) the strong central band has no counterpart in the spectrum of ribonuclease A which, by an older terminology, would be said to have ca. 1/3 "random" structure. The "random" portions of the ribonuclease chain are thus by no means similar to a denatured protein. The latter most probably has its backbone carbonyls hydrogen bonded to water molecules (2), not to other peptide groups.

The remaining spectral features are not as easily interpreted and are not related to the secondary structure. The weak peak around  $1600\text{ cm}^{-1}$  is too low for an amide I' component (1,2); it is probably caused by aromatic side-chain groups (17,18). Other sidechain groups of histidine, tryptophan, and phenylalanine, as well as the asymmetric stretching mode of sidechain  $\text{COO}^-$  groups also absorb in the  $1550$  to  $1610\text{ cm}^{-1}$  region (17,18). The very stable, sharp  $1515\text{ cm}^{-1}$  band is probably associated with tyrosine residues (18). From ca.  $1430$  to  $1480\text{ cm}^{-1}$  we have overlapping bands caused by (a)  $\text{CH}_2$  and  $\text{CH}_3$  bending modes of side chains (17); (b) the amide II' mode, essentially ND bending (2,4); and (c) bending modes of traces of HOD (2,4). From  $1360$ - $1380\text{ cm}^{-1}$  the symmetric  $\text{CH}_3$  bending vibrations of sidechains are expected (17). More detailed assignments must wait a thorough study of more proteins. It is evident, nevertheless, that second derivative spectra furnish new information about the sidechains as well as the secondary structure of proteins.

#### ACKNOWLEDGEMENT

The authors thank Janine Brouillette for her invaluable assistance in preparing solutions and obtaining the excellent infrared spectra.

#### REFERENCES

1. Krimm, S. (1962) J. Mol. Biol. 4, 528-540.
2. Susi, H., (1972) Meth. Enzymol. 26, 455-472.
3. Timasheff, S. N., Susi, H., and Stevens, L. (1967) J. Biol. Chem. 242, 5467-5473.
4. Ruegg, M., Metzger, V., and Susi, H. (1975) Biopolymers 14, 1465-1471.
5. Lippert, J. L., Tyminski, D., and Desmeules, P. J. (1975) J. Am. Chem. Soc. 98, 7075-7080.
6. Williams, R. W. and Dunker, A. K. (1981) J. Mol. Biol. 152, 783-813.
7. Richardson, J. S. (1981) Adv. Protein Chem. 34, 167-363.
8. Maddams, W. F. and Tooke, P. B. (1982) J. Macromol. Sci. A17, 951-968.

9. Savitzky, A. and Golay, M. J. E., (1964) *Anal. Chem.* 36, 1627-1639.
10. Kauppinen, J. K., Moffat, D. J., Mantsch, H. H., and Cameron, D. G. (1981) *Anal. Chem.* 53, 1454-1457.
11. Maddams, W. F. and Southon, M. J. (1982) *Spectrochim. Acta* 38A, 459-466.
12. Levitt, M. and Greer, J. (1977) *J. Mol. Biol.* 114, 181-293.
13. Kawai, M. and Fasman, G. (1978) *J. Am. Chem. Soc.* 100, 3630-3632.
14. Bandekar, J. and Krimm, S. (1979). *Proc. Natl. Acad. Sci. U.S.A.* 76, 774-777.
15. Timasheff, S. N., Townend, R., and Mescanti, L. (1966) *J. Biol. Chem.* 241, 1863-1870.
16. Susi, H. and Byler, D. M., unpublished infrared spectra.
17. Bellamy, L. J. (1975) *The Infra-red Spectra of Complex Molecules*, Vol. 1, 3rd ed., pp. 6, 8, 21-27, 198-200, Chapman and Hall, London.
18. Chirgadze, Yu. N., Fedorov, O. V., and Trushina, N. P., (1975) *Biopolymers* 14, 679-694.